WILEY InterScience® DISCOVER SOMETHING GREAT

A Bibliometric Model for Journal Discarding Policy at Academic Libraries

Evaristo Jiménez-Contreras, Mercedes De La Moneda, and Elvira Ruiz de Osma
Facultad de Documentación, Campus de Cartuja, Universidad de Granada, 18071-Granada, España. E-mail: {evaristo, dlmoneda, elruiz}@ugr.es

Rafael Bailón-Moreno
Departamento de Ingeniería Química, Facultad de Ciencias, Campus de Fuentenueva, Universidad de Granada, 18071-Granada, España. E-mail: bailonm@ugr.es

Rosario Ruiz-Baños
Facultad de Documentación, Campus de Cartuja, Universidad de Granada, 18071-Granada, España. E-mail: rruizb@ugr.es

The authors propose a bibliometric model for discarding journal volumes at academic libraries, i.e., removal to offsite storage as part of the library's serials collection. The method is based on the volume as the unit of measurement and on user satisfaction with given titles. The discarding age, calculated for each volume, from the year of publication to the year of decision to discard, is dependent on citation half-life, relative productivity, knowledge area, and residual utility (potential consultations). The model makes it possible to predict the approximate size of a collection when a stationary state is reached in which the inflow of journal volumes is equal to the outflow from discarding. The model is also able to determine the rate of growth of the holdings. This information can be used to optimize future use of available space and economic and maintenance resources; thus promoting efficient management of the collection.

Introduction

Discarding—defined here as the removal from the shelves of part of a library's serials collection—is a difficult undertaking. If not based on pertinent criteria, it can prevent access by users to useful, profitable information. Because of the importance of this procedure, many studies have attempted to identify which factors and variables should be brought to bear in the design of an efficient discarding policy.

Lancaster (1993) has reviewed different discarding methods, involving criteria such as current use of a collection, citation and impact factors, obsolescence, expert opinion, pertinence, and naturally costs. Much debate now centers on the specific weight of internal data (usage, costs, etc.) versus external data (information from other libraries and nonlibrary sources) in discarding criteria. Segal noted that because the number of variables considered significant is constantly increasing, each should be assigned a specific weighting and different factors should be combined in mathematical formulas. The weighting criteria proposed by Segal (1986) assigns 5 points each for (a) a complete collection; (b) if the cost of the subscription to a given journal is equal to or lower than the mean cost of subscriptions to other journals that cover the same material; (c) the impact factor is equal to or higher than the mean impact factor of other titles that cover the same subject matter; (d) if the title is cited as fundamental in the bibliography compiled by Katz and Katz (1997); and (e) if the subject matter is taught at the center the library serves. If the publication is indexed by all five of the main databases that specialize in the material, 1 to 5 points are assigned.

Criteria for canceling journal subscriptions, such as those proposed by Hunt (1990), may be of use in designing quantitative criteria for discarding. This author proposed a formula for comparing the cost of maintaining a subscription with the cost of providing access via interlibrary loan.

$$R = \frac{CI}{P + M + LS} \tag{1}$$

where R is the institutional cost ratio; C is the annual number of consultations; I is the price per interlibrary loan request; P is the price of subscriptions; M is the annual cost of subscription maintenance; L is the number of linear meters of shelf space occupied by bound volumes in the collection; and S is the cost of storage per meter of shelf space.

Received June 28, 2004; revised November 23, 2004; accepted November 23, 2004

 • Published online 16 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20264

The variable C—the number of consultations per year—is an important element to consider in relation to total annual storage costs (LS) when discarding policies are being developed. Although different methods have been proposed to quantify this variable, usage of the collection is a common factor in all discarding policies. Of note are those methods that are based on internal data to measure (for example) journal usage in the reading room, and those that investigate usefulness of the collection for patrons and researchers with publications available in the literature by counting citations to their articles (Jiménez-Contreras, Moneda, Overa, & Ruiz de Osma Delatas, 1994). Other methods use external data to determine (for example) the potential use of the collection by library patrons, and extrapolate the results obtained in the study of sources cited to a similar but larger community. To define the user community under study, one option is to choose the community of researchers who publish in journals to which the library subscribes. In principle, the adequate degree of similarity between these two research communities is ensured by the use of similar information sources and the same disciplines.

The present study was designed to evaluate the usefulness of a collection with citation analysis, a tool managers need to consider when they develop discarding policies. We do not intend for the approach reported here to replace all other alternatives, nor has it been fully tested and refined. We simply present a quantitative approach to a complex issue. The methods described are, to a large extent, original, and are derived from earlier work by Curtis (1975), who proposed a predictive model based on the use of ageing as estimated from citations to "weed" journals and satisfy users' needs. Slote (1997) also proposed utility as an approach to weeding. This author suggested a user satisfaction criterion of 90 to 95%, and noted that the items representing the remaining 5% or 10% of the requests could be considered dispensable. However, Slote did not attempt further quantitative analysis; his opinion was that mathematical methods can not be used effectively with ideas he considered imprecise.

The problem, which we attempt to deal with in this article, is to establish a quantitative criterion that will make it possible to determine what proportion of the collection serves users' needs. In this connection, bibliometrics offers an approach based on Brookes' formula for obsolescence (termed the "aging factor" in his study), that provides the theoretical and quantitative underpinning for our method (Brookes, 1970).

Brookes defines total or initial utility, $U(0)$, of a recently published journal volume with an infinite life expectancy, as the total number of citations expected from the date of publication. Utility decreases with time according to a negative exponential function, such that its initially maximal value $U(0)$ decreases and eventually becomes zero after a time that approaches infinity. In mathematical terms, this formula is expressed by the equation:

$$U(t) = U(0)a^t \tag{2}$$

The annual aging factor, a, can range in value from 0 to 1. When $a = 0$, aging is instantaneous, and when $a = 1$, aging does not occur. The aging factor is related with half-life, h, through the expression:

$$a^h = \frac{1}{2} \tag{3}$$

As a result, aging factor and half-life refer to essentially equivalent concepts. Values of $a > 0.9$ ($h > 6.7$ years) are obtained for journals of high currency, that is, journals that are highly relevant to researchers' needs and which therefore age slowly. Values between 0.8 and 0.9 (h between 3.1 and 6.7 years) can be considered to reflect the usual pattern of aging, and values of $a < 0.8$ ($h < 3.1$ years) indicates that the information is ephemeral.

Burton and Kebler (1960)—the originators of the concept of citation half-life—noted that this parameter varies between disciplines. Brookes described two main reasons for consulting a journal volume: immediate and historical interest. Items with immediate interest age more rapidly than items sought for their historical interest. Griffith and colleagues later measured the aging rate of the entire body of literature cited in the *Science Citation Index* (*SCI*; Thomson ISI, Philadelphia, PA), and confirmed the phenomenon described by Brookes (Griffith, Servi, Anker, & Drot, 1979). Terrada and colleagues studied a sample of biomedical journals and noted that the aging rate depended on the country of origin of the article and the language of publication (Terrada, Cueva, & Añon, 1979). Recently, Ruiz-Baños and Jimenez-Contreras (1996) found that in a sample of library and information science journals, those that formed part of the "advancing scientific front" (i.e., disciplines in which knowledge, and therefore publications, are increasing rapidly) aged more slowly than journals on the periphery. According to Zhao and Jiang (1985), the predominant scientific front comprises the USA, the UK, and English-speaking countries in general. Journals published in these countries are characterized by their use of English almost exclusively, by high impact factors, and by good coverage in *SCI*.

An alternative to Brookes' model was recently proposed (Álvarez, Escalona, & Pulgarin, 2000). The alternative is also based on citation analysis, however, instead of a simple exponential model such as that shown in Equation 2, it proposes a probabilistic model as put forward Rasch (1980).

The probability that journal n will be cited in year i is:

$$P\{X_{ni} = 1/\beta_n, \delta_i\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \tag{4}$$

where β_n represents journal n and δ_i represents year i.

The model proposed by Rasch and applied to journal selection procedures by Alvarez et al. is promising but is hampered by the complexities of determining β_n and δ_i. In addition, the mathematical expression is more complex than that used by Brookes, and subsequent calculations lead to highly complex expressions that are difficult to use in practice.

DOI: 10.1002/asi

We therefore preferred to use the classical expression proposed by Brookes to calculate the discard age for a given journal volume with the following expression (Ruiz-Baños, 1994):

$$t_e = \frac{\ln\left(p\frac{j}{i}\right)}{\ln a} \tag{5}$$

where t_e is discard age in years; p is the decimal fraction of global residual utility; j is the mean number of relevant articles per volume for the entire collection at the library; i is the number of relevant articles in the volume under consideration; and a is the annual aging factor.

Based on the annual aging factor a and the number of relevant articles i, Ruiz-Baños established three hypothetical cases for university libraries. In the ideal (and most unlikely) scenario, a is constant and i is the same for all journals. In the second scenario, the aging rate a is constant and i differs between journals. This may be the case when all journals held by the library are similar in terms of obsolescence. The third and most likely scenario assumes that journals differ widely in obsolescence (a) and in the number of relevant articles (i) (Ruiz-Baños, 1994). Many academic libraries are facing this situation.

Aim of the Study

In theoretical terms, the concepts of annual aging factor (a) and half-life (h) are equivalent, as shown in Equation 3. Because half-life is more intuitive and is more widely used, we rewrote Equation 5 with h as the main factor. This yields a more manageable expression that we use below to develop a model to predict how a journal collection evolves with time. The resulting model should be able to determine: (a) a reasonable discard age for each volume of each journal, (b) the maximum size of the collection when stationary state is reached, in which the number of incoming volumes equals the number of volumes removed from the shelves through discarding, and (c) the growth rate of the collection.

Background to the Model

As shown above, the model Brookes proposed evaluates the decrease in utility of science journals with time. According to this author, a volume's utility is determined by whether it has been used as a source for the writing of subsequent scientific documents such as journal articles or congress presentations. Its use as a source can be quantified as the number of times it is cited by the rest of the scientific community.

$$\textit{Utility} \Leftrightarrow \textit{Use} \Leftrightarrow \textit{Citations Received} \tag{6}$$

Two widely used indicators of citation are the impact factor (IF) and the citation half-life (h). The former is defined as the mean number of citations received per article in a given volume during the 2 years following publication. As its name suggests, the IF reflects the initial impact of the volume in the scientific community, and is therefore appropriate for decisions regarding acquisitions, choice of journal for manuscript submittal, and research evaluation, among other issues. In contrast, and as we will show below, it seems reasonable to consider the IF unsuitable as an indicator of behaviors such as discarding, which requires a period of observation longer than 2 years.

Citation half-life is the period that elapses between publication and the appearance of half of the total number of citations the volume eventually accumulates. This period can be much longer than the 2-year window used to calculate the IF. Accordingly, citation half-life reflects the currency of the volume. A volume cited during many years after publication has a long half-life, which implies greater interest by researchers, greater use, and hence greater utility.

The changes in the number of citations a volume receives follow a characteristic pattern, with some variations. In many cases, citations increase steadily and peak after about 2 years, then decrease exponentially and eventually cease altogether. However, peak citation rates can appear as long as 5 or 6 years after publication. The rate of decline in citations can also vary. Figure 1 shows four curves illustrating variations in citation behavior. This figure shows that IF is not necessarily related to currency.

To further illustrate this, Figure 2 plots the IF against half-life for some journals indexed in the *Journal Citation Reports* (*JCR*; Thomson ISI, Philadelphia, PA) in the areas of biology, physics, geology, mathematics, and chemistry for the year 1997. The lack of association between IF and half-life is evident, with a R^2 of practically zero. In other words, this figure shows empirically that the IF is not a useful predictor of the subsequent use of a given volume of a journal. In contrast to other authors, and in light of the data from the *JCR*, we feel that the IF is not a useful factor for decisions to discard, which are made more than 2 years after publication. In the subsequent description and discussion of our model for discarding, therefore, no further mention is made of the IF. Instead, we have opted to use citation half-life as our starting point.

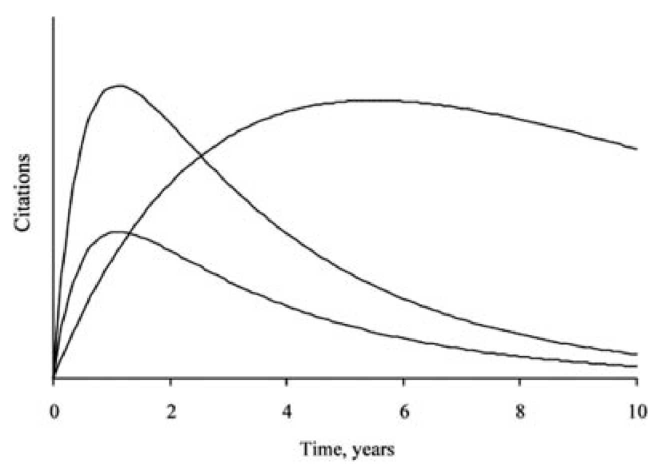

FIG. 1. Citations to articles in journals.

DOI: 10.1002/asi

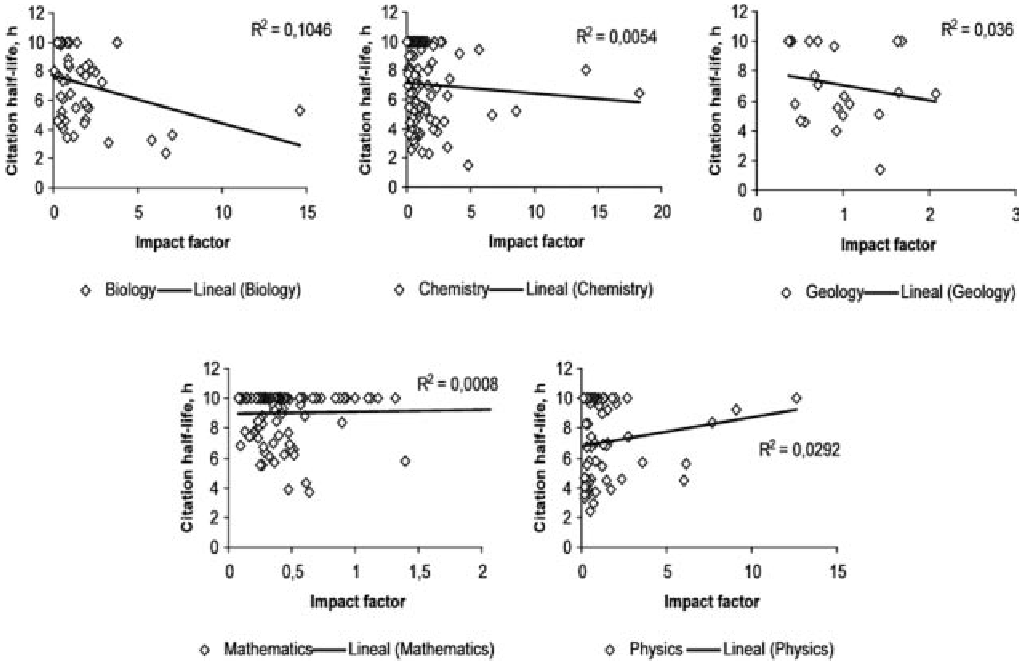

FIG. 2. Influence of citation half-life on impact factor in five knowledge areas.

Discard Age

Consider a library with holdings that cover J disciplines, each of which is represented yearly by a mean productivity of n_j articles in $I(j)$ volumes. For example, let us consider that volume i in discipline j comprises $n_{i,j}$ items or articles. Mean productivity of the area, n_j, is the quotient obtained by dividing the sum of all items published in all volumes by the total number of volumes in the area:

$$\bar{n}_j = \frac{\sum_{i=1}^{I(j)} n_{i,j}}{I(j)} \tag{7}$$

According to Equation 3, for a volume with half-life $h_{i,j}$, the expression can be recast as:

$$\frac{1}{2} = a_{i,j}^{h_{i,j}} \tag{8}$$

Taking the logarithm for terms on both sides of the equation yields:

$$\ln\frac{1}{2} = h_{i,j}\ln a_{i,j} \tag{9}$$

After solving, the logarithm of the annual aging factor for volume i,j is:

$$\ln a_{i,j} = -\frac{\ln 2}{h_{i,j}} \tag{10}$$

From Equation 5, Equation 10 can be recast as:

$$t_{i,j} = \frac{\ln\left(p\dfrac{\bar{n}_j}{n_{i,j}}\right)}{\ln a_{i,j}} \tag{11}$$

Therefore discard time $t_{i,j}$ can be expressed as:

$$t_{i,j} = \frac{\ln\left(p\dfrac{\bar{n}_j}{n_{i,j}}\right)}{-\dfrac{\ln 2}{h_{i,j}}} \tag{12}$$

Solving for ln(2) and moving $h_{i,j}$ to the numerator yields:

$$t_{i,j} = -1.44h_{i,j}\ln\left(p\frac{\bar{n}_j}{n_{i,j}}\right) \tag{13}$$

By developing the logarithm with care not to change the signs, discard time can be expressed as:

$$t_{i,j} = 1.44h_{i,j}\left[\ln\frac{n_{i,j}}{\bar{n}_j} - \ln p\right] \tag{14}$$

According to this equation, discard time is directly proportional to citation half-life, and is also directly proportional to a complex function defined by the logarithm of relative productivity of the volume and the cologarithm of residual utility.

Relative productivity, $r_{i,j}$, is defined as the quotient of productivity of the volume divided by mean productivity of the area:

$$r_{i,j} = \frac{n_{i,j}}{\bar{n}_j} \tag{15}$$

Equation 14 can also be written as follows:

$$t_{i,j} = 1.44h_{i,j}\lfloor \ln r_{i,j} - \ln p \rfloor \tag{16}$$

DOI: 10.1002/asi

To match calculated discard time as closely as possible to the needs of the community, we recommend using a half-life value calculated from citations by library users. The values provided by the *Journal Citation Reports* can be used if the community of library users is assumed to be homogeneous to the international community.

One issue that deserves attention is endogamy, which can appear as a result of the half-life calculated from citations by the local community. This might lead to discarding of journals that are widely accepted by the international community, but whose usefulness is overlooked by local users. For such cases, librarians should use the *JCR* half-lives, which are higher. This would ensure that valuable journals remain available to readers, who may obtain new insights and perspectives from their contents.

To summarize, we recommend using the $h_{i,j}$ value calculated for citations by library users as long as this figure is approximately equal to or higher than the *JCR* half-life. If the users' half-life is very much shorter than the *JCR* figure, and if the latter is very long, the latter should be used to create incentives to use the journal(s) under consideration. If the two figures are similar, the local figure is preferable.

Residual utility p represents potential use, measured in terms of relative citations that a discarded volume might have received if it had been left on the shelves. This factor is determined by the librarian's decision, and is related with the chosen level of user satisfaction the librarian wishes to satisfy. According to Slote, a less stringent criterion is a 10% rate of user "insatisfaction" (i.e., 90% of all users would be satisfied during the period under study), and a more stringent criterion in 5% (95% of all users satisfied). Because p is expressed as a decimal fraction, its value can be 0.10 or 0.05. This factor is independent of the knowledge area and the number of volumes considered, and depends instead on user satisfaction policies in effect at the library. Below we will suggest some objective criteria for choosing the most appropriate value of residual utility. These criteria are related with factors such as acquisition policy, available space, and costs.

Discard Time as a Function of Half-Life

Whereas productivity and residual utility are expressed as a logarithm, discard time is directly proportional to half-life. Therefore, half-life, which represents the real use of a given volume, is the variable with the greatest relative weight. However, the slope of the line obtained by plotting t_{ij} against h_{ij} depends on productivity and residual utility according to the equation:

$$m = 1.44(\ln r_{i,j} - \ln p) \qquad (17)$$

Thus, relative weight of the half-life is affected by productivity and residual utility. The slope increases as r_{ij} increases and as residual utility decreases. In other words, utility of the collection, h_{ij}, is greater for highly productive volumes and for higher user satisfaction criteria.

Figure 3 shows that for a given residual utility, the slope of the line increases with relative productivity. Similarly, for a given productivity value, discard time increases as residual utility decreases. This is illustrated by the higher slopes in Figure 3B in comparison to Figure 3A.

Discard Time as a Function of Productivity

If all other variables are held constant, discard time is proportional to the logarithm of productivity. This means that a small increase in a low productivity will lead to a large increase in discard time. In contrast, when productivity is high, discard time will increase noticeably only with much larger increases in productivity.

Figure 4 shows the results of plotting $t_{i,j}$ against $n_{i,j}$. The lines form two bundles: the upper one reflects volumes with a half-life more than threefold as long as that of the volumes represented in the lower group. Within each bundle, each line illustrates the influence of mean productivity of the knowledge area, n_j. In general, discard time for a given journal is more sensitive to changes in use habits for that journal than to changes in productivity of the discipline to which

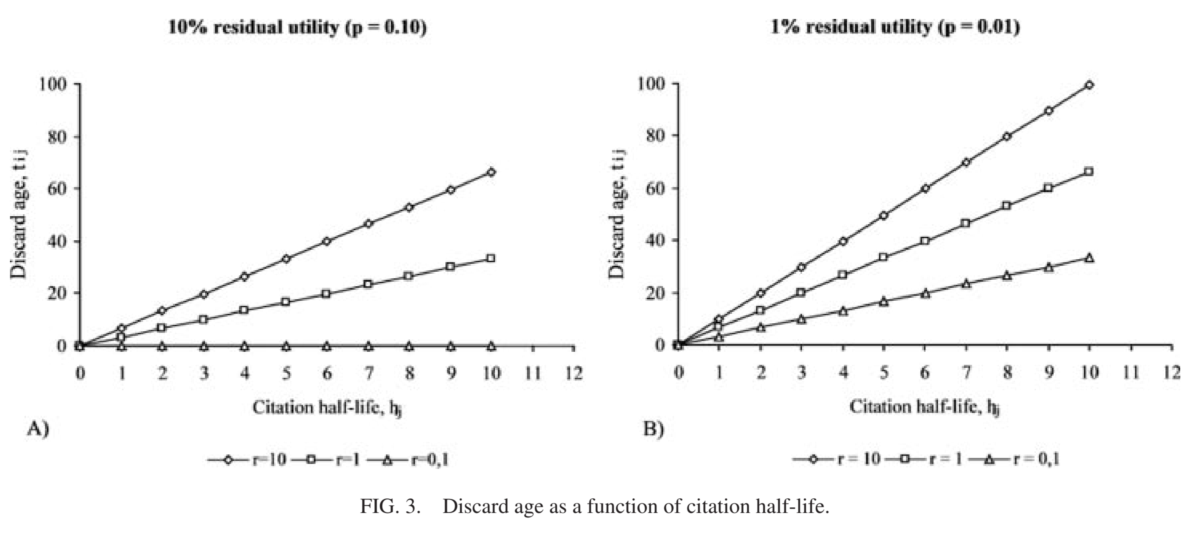

FIG. 3. Discard age as a function of citation half-life.

DOI: 10.1002/asi

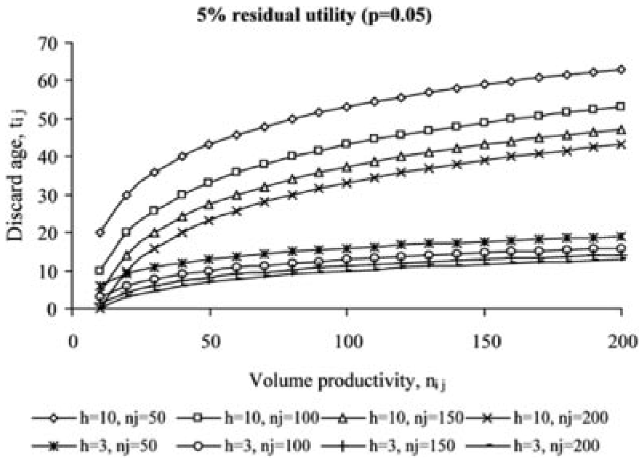

FIG. 4. Discard age as a function of productivity.

they belong. The influence of discipline is even weaker for journals with a short half-life.

Discard Time as a Function of Residual Utility

Because t_e is proportional to the cologarithm of residual utility, the curve obtained when t_e is plotted against this variable shows decreasing values as the latter increases (Figure 5). It follows that if a journal is deselected too soon, the residual utility, i.e., lost opportunities to obtain useful information from the journal, increases, and user satisfaction decreases. This again reflects the usefulness of use of the collection, measured as citation half-life, for determining discard time, and also illustrates the influence of use on the slope of the resulting curves.

The Maximum Residual Utility

Turning once more to Figure 3A, it is interesting to note that in the line for $r = 0.1$, discard time is 0 regardless of the half-life of the volume in question. The resulting slope can in fact be 0 or have a negative value:

$$1.44(\ln r_{ij} - \ln p) \leq 0 \tag{18}$$

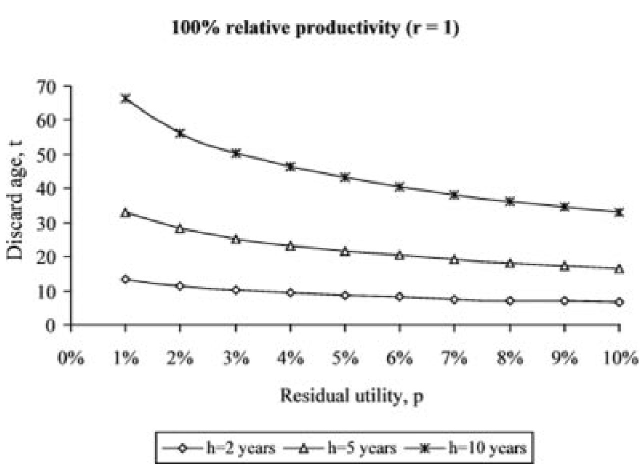

FIG. 5. Discard age as a function of residual utility.

After solving, the resulting expression is:

$$\ln r_{ij} \leq \ln p \tag{19}$$

$$p \geq r_{ij} \tag{20}$$

In other words, if discard utility, understood here to mean the residual utility of the volumes to be removed, were greater than relative productivity of the volume, the formula would yield a discard time of 0 or even a negative value. This means that if too high a p value is used, some volumes would be deselected within a year of their acquisition—the implication being that they should not have been added to the collection in the first place. This apparent paradox provides a clue as to which quantitative criteria librarians should use to maintain a minimally acceptable level of user satisfaction, without violating acquisition and cancellation policies. The residual utility chosen as the criterion should be sufficient to ensure that the least productive journal is discarded no sooner than a year after publication ($t_{ij} \leq 0$). This avoids a conflict between criteria based on factors such as suitability, excellence, or impact factor, used to select titles for acquisition, user satisfaction criteria, and discarding policies. In summary, *residual utility, p, should never be greater than the lowest relative productivity,* r_{ij}. The criterion illustrated in Figure 3B is thus more reasonable than the criterion in Figure 3A. If p is too high (e.g., 0.10), the result would be that journals with low relative productivity would be removed too soon regardless of their half-life.

The Stationary State and Collection Size

By stationary state of a library, we mean the state at which the size of the collection remains constant. This occurs when the incorporation of new volumes is offset by the loss, through discarding, of other volumes.

$$\{\textit{Inflow from new volumes added}\} = \{\textit{Outflow from old volumes discarded}\} \tag{21}$$

With Equation 14 or its equivalent, Equation 16, we can calculate the maximum size of a collection. To simplify the calculations, a number of assumptions can be used:

1. The publication policies of journals are assumed to be constant in the sense that a given journal continues to publish about the same number of articles (available items) per year, as described by Griffith et al. (1979). If this is not the case, increased productivity by some journals can be assumed to be offset by similar decreases in productivity in others.
2. There are no substantial changes in users' citation behavior during the period of study. In other words, the half-life remains more or less constant, as described by Griffith et al. (1979).
3. As a result, all volumes of a given journal can be assumed to have the same discard time.
4. No new titles are acquired and there are no cancellations.

DOI: 10.1002/asi

A more general model with fewer restrictions will be described in a later publication.

The Case of a Single Journal Title

We will first analyze a simple case in which a collection grows in size through the addition of a single new title. During the first year, only one volume will be added to the collection; during the second year a second volume will be added, and so on. Volumes accumulate yearly until the year when the journal is deselected. In that year the entering volume takes the place, so to speak, of the oldest volume acquired, which is discarded. Thereafter, a stationary state ensues: The size of the collection remains constant at the size it had reached in the year the journal was discarded. *The number of volumes of the journal is equal to the number of years to discarding since the title was first added to the collection.* This is expressed as:

$$v_{i,j} = t_{i,j} \tag{22}$$

where $v_{i,j}$ is the number of volumes of title i (belonging to area j), and $t_{i,j}$ is the discard time.

The Case of Several Titles of the Same Discipline

We now consider area j, represented by $I(j)$ titles. In this case, the number of volumes in stationary state is equal to the sum of all volumes of each title:

$$v_j = \sum_{i=1}^{I(j)} v_{i,j} \tag{23}$$

Equation 22 and Equation 14 yield:

$$v_j = \sum_{i=1}^{I(j)} t_{i,j} \tag{24}$$

$$v_j = \sum_{i=1}^{I(j)} \left[1.44 h_{i,j} \left(\ln \frac{n_{i,j}}{\overline{n}_j} - \ln p \right) \right] \tag{25}$$

$$v_j = \sum_{i=1}^{I(j)} \left[1.44 h_{i,j} \ln \frac{n_{i,j}}{\overline{n}_j} - 1.44 h_{i,j} \ln p \right] \tag{26}$$

$$v_j = 1.44 \sum_{i=1}^{I(j)} \left(h_{i,j} \ln \frac{n_{i,j}}{\overline{n}_j} \right) - 1.44 \sum_{i=1}^{I(j)} [h_{i,j}] \ln p \tag{27}$$

$$v_j = a_j - b_j \ln p \tag{28}$$

where the constants a_j and b_j are defined, respectively, as

$$a_j = 1.44 \sum_{i=1}^{I(j)} \left(h_{i,j} \ln \frac{n_{i,j}}{\overline{n}_j} \right) \tag{29}$$

and

$$b_j = 1.44 \sum_{i=1}^{I(j)} h_{i,j} \tag{30}$$

As seen from Equation 28, the maximum size of the collection is a function of the logarithm of residual utility chosen by the librarian. If a high p value (lower user satisfaction) is used, smaller collections (which are less costly to maintain) will result. On the other hand, if a low p value is used, size of the collection and hence user satisfaction will be greater, but so will the maintenance costs.

The Case of an Entire Collection

Here we assume a collection consisting of J knowledge areas, each of which behaves differently. If the maximum number of titles in each area, j, is equal to v_j, the maximum size of the collection, V, can be calculated from the sum of the volumes held for all areas:

Replacing v_j with $(a_j - b_j \ln p)$ yields

$$V = \sum_{j=1}^{J} (a_j - b_j \ln p) \tag{31}$$

Developing the sum yields:

$$V = \sum_{j=1}^{J} a_j - \sum_{j=1}^{J} (b_j) \ln p \tag{32}$$

The size of the entire collection is calculated as:

$$V = \sum_{j=1}^{J} v_j \tag{33}$$

$$V = a - b \ln p \tag{34}$$

where the new constants a and b are expressed as

$$a = \sum_{j=1}^{J} a_j \tag{35}$$

$$b = \sum_{j=1}^{J} b_j \tag{36}$$

Equation 34 is similar to Equation 29 for size of the collection in a given knowledge area; hence, the size of the collection depends on the value used for residual utility. For high p values and low levels of user satisfaction, the results point to small collections that are less expensive to maintain. In contrast, low p values yield high levels of user satisfaction with large collections, which use more space and are more costly to maintain.

Above we showed that the highest possible value of p was the greatest value compatible with the acquisitions policy. The criterion then becomes to find the value of p equal to or higher than the lowest relative productivity, $r_{i,j}$, for the entire collection. Equation 34 makes it possible to choose a given cost or a given amount of physical space to determine the minimum residual utility that can be used to develop the discarding policy. *This value of minimum usefulness should be chosen to*

DOI: 10.1002/asi

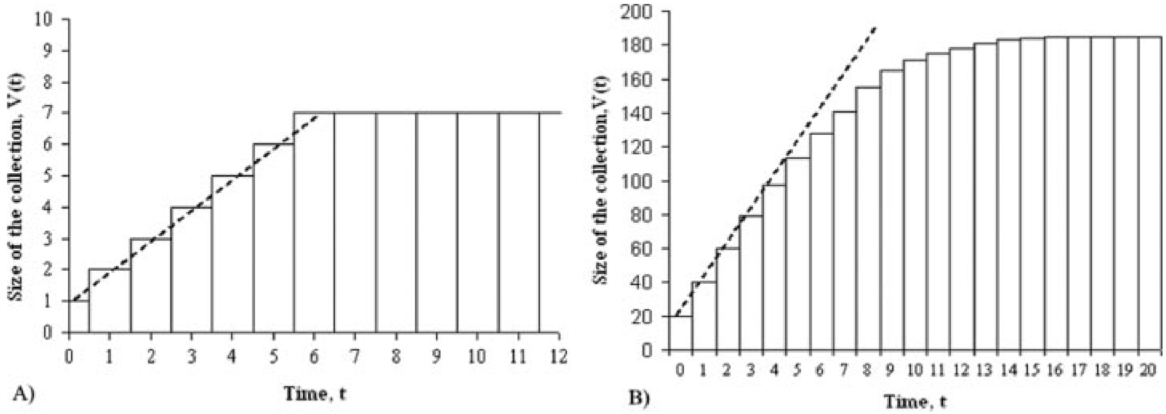

FIG. 6. Growth rate of a collection, shown for a single title (A) and for a group of titles, a knowledge area, or a complete collection (B).

ensure that the stationary state generates a collection that will fit on the library's available shelf space, and therefore will not overrun the maximum budget available for maintenance costs. Solving for p in Equation 34 yields the minimum possible value, p_{min}, for a collection whose size is $V = f\,V_{max}$:

$$p_{min} = e^{\frac{a - fV_{max}}{b}} \tag{37}$$

$$p_{min} = Ae^{-\frac{fV_{max}}{b}} \tag{38}$$

where V_{max} is the maximum carrying capacity of the library expressed in volumes, f is the recommended rate of shelf occupancy, allowing for a margin of error to accommodate possible extra volumes, and the constant A is expressed as

$$A = e^{-\frac{a}{b}} \tag{39}$$

Minimum utility is a negative exponential function that involves an infinitely long asymptotic tail that approaches 0. The value of p would be null, and user satisfaction would be 100%, only for an infinitely large collection of journals. As this is untenable, we conclude that complete user satisfaction is impossible to attain. Therefore, *by establishing the size of the collection on the basis of available space and available maintenance budget, we can determine in advance the associated residual utility and the degree of user satisfaction attainable.*

Growth Rate of the Collection

Once the target for residual utility has been established, and once discard time has been calculated for each volume, it becomes possible to determine the rate at which the collection can be expected to grow. This is useful because it makes it possible to predict the likely space and staff needs with time.

The Case of a Single Journal Title

For title i belonging to area j, and assuming the discard time is not exceeded, the number of volumes $v_{i,j}(t)$ of the title for each year t, will be equal to the number of volumes from the preceding year plus 1. In other words, $v_{i,j}(t)$ will be equal to the number of years the title has been carried. When the discard age is reached, a stationary state ensues and the number of volumes remains constant.

$$\begin{aligned} v_{i,j}(t) &= v_{i,j}(t-1) + 1 \qquad \forall\, t \le t_{i,j} \\ v_{i,j}(t) &= v_{i,j}(t_{i,j}) = t_{i,j} \qquad \forall\, t > t_{i,j} \end{aligned} \tag{40}$$

Figure 6A shows how the number of volumes of a given title increases as the size of the collection increases until time t_e, which marks the end of the period of growth.

The Case of a Group of Journals

This example looks at the case of $I(j)$ titles belonging to area j. During the first year $I(j)$ volumes will enter the collection, one for each title. At the end of the second year the total number of volumes will be $2I(j)$, and so on until the year when discarding begins. Thereafter, the growth rate will be $I(j) - 1$ volumes per year. After the second year of discarding, the growth rate will be $I(j) - 2$, and so on until the oldest volume of a given journal in the collection is to be deselected. From that time on, annual growth is zero and stationary state ensues. In mathematical terms, the size of the collection, expressed as the number of volumes, v_j, in year t is:

$$v_j = v_j(t-1) + c(t) \tag{41}$$

where $c(t)$ is annual growth at a given time.

$$\begin{aligned} t = 1 \qquad & c(t) = I(j) \\ t = t_{i,j}\ (greatest) \qquad & c(t) = 0 \end{aligned} \tag{42}$$

The difference between this case for several journals and the case for a single journal is only that in the former, annual growth decreases steadily. In the latter case, however, growth is constant until the time for discarding is reached; at which point growth ceases abruptly. This difference is illustrated in Figures 6A and 6B.

Growth rate, $c(t)$, for a given year is equal to the preceding year's growth rate minus the number of titles that have reached their discard age. Hence:

$$c(t) = c(t-1) - e(t-1) \tag{43}$$

where $e(t-1)$ represents the number of titles that have reached discard age, $t_{i,j}$, in year $t-1$. The expressions shown in Equation 42 are the limiting conditions for Equation 43.

Summary

To recapitulate, the model we propose is based on the relationship between the concept of utility, as described by Brookes (1970), and user satisfaction.

The utility of a journal volume is considered in the light of actual use as a source for the preparation of new scientific documents. The most suitable variable to quantify actual use of a given item is the number of citations received.

The expression we developed to calculate discard age of a given volume *i* belonging to discipline *j* is:

$$t_{i,j} = 1.44 h_{i,j}\left[\ln\frac{n_{i,j}}{\bar{n}_j} - \ln p\right] \tag{44}$$

Discard age is directly proportional to the half-life, $h_{i,j}$, of the citations received, and to a complex function derived from the logarithm of relative productivity, $n_{i,j}/n_j$, and cologarithm of residual utility, p, which the librarian assigns to the collection.

In short, to determine the time of discarding of a particular volume, our model uses mainly both journal-intrinsic factors, tied to its publishing policy, and journal-extrinsic factors of a library and scientific nature.

Residual utility, hence, is related with user satisfaction. Ideally, residual utility should be chosen to lie between two other values: The upper limit should not exceed the lowest relative productivity, $r_{i,j}$, of any journal in the collection. The lower limit is determined by the expression:

$$p_{min} = Ae^{-\frac{fV_{max}}{b}} \tag{45}$$

The minimum residual utility, p_{min}, that the collection must have relates user satisfaction with the physical space available in the library and with the maximum estimated maintenance costs. It is necessary to point out that the complete satisfaction, $p_{min} = 0$, is not possible, because it would involve having an infinitely large collection. In addition, the growth of costs becomes exponential.

A stationary state is said to have been reached when the inflow of new volumes is equal to the outflow of deselected volumes. Under these conditions, the size of the collection remains constant, although the final size will vary depending on the residual utility policy used by collection managers. The size of the collection is calculated with the equation:

$$V = a - b \ln p \tag{46}$$

V is the size in terms of volumes, which is easily transformable in shelving length or floor space, by applying a proportionality factor.

By analyzing the dynamics of a collection, it becomes possible to calculate its growth rate for the period until stationary state is reached. The growth kinetics can be described by the set of equations below:

$$v_j = v_j(t-1) + c(t) \tag{47}$$

$$\begin{matrix} t = 1 & c(t) = I(j) \\ t = t_{i,j}(greatest) & c(t) = 0 \end{matrix} \tag{48}$$

$$c(t) = c(t-1) - e(t-1) \tag{49}$$

In the hypothetical case of only having one copy of a journal, growth is lineal or constant until it levels out and suddenly becomes nil. When a group of titles is considered however, which is the real case, the growth of the collection is not lineal. It is at its highest during the first year and then decreases progressively. This is because there are more and more journals reaching their stipulated withdrawal date. Therefore, while the complete collection reaches its overall stationary level, each individual journal within the collection reaches stationary micro-levels. This means that the space occupied is going to be filled very quickly during the first few years, stabilizing progressively over the following years.

As a result of applying our new model, a librarian knows beforehand the future behavior of the serials library as far as its growth and size are concerned.

Further Developments of Our Model

The model may be developed further to take into account specific collections and changes in usage, by considering new acquisitions and cancellations, changes in users' citation behavior, and changes in journals' publication policies. Approaches to the extension of the model will be reported in a separate publication. Efforts to develop a PC-based application to facilitate implementation of the discarding procedure are also underway.

Ideally, the model should take into account the influence of the Internet and electronic access on user behavior, as electronic editions of journals are replacing printed editions as sources of information. As a result, the model will incorporate terms for the use of online journals, for example, that bandwidth or the channel capacity of transmission is limited. The increasing use of electronic journals may lead to premature discarding or even cancellation of subscriptions to printed journals.

References

Alvarez, P., Escalona, I., & Pulgarín, A. (2000). What is wrong with obsolescence? Journal of the American Society for Information Science 51, 812–815.

Brookes, B.C. (1970). Obsolescence of special library periodicals: Sampling errors and utility contours. Journal of the American Society for Information Science, 21, 320–329.

Burton, R.E., & Kebler, R.W. (1960). "The half-life" of some scientific and technical literatures. American Documentation, 11, 18–22.
Curtis, D.T. (1975). A predictive model to cost-control older scientific materials in libraries. In C.W. Husbands & R.L. Tighe (Eds.), Proceedings of the 38th ASIS Annual Meeting (pp. 49–50). Washington, DC: American Society for Information Science.
Griffith, B.C., Servi, N.P., Anker, A.L., & Drot, M.C. (1979). The aging of scientific literature: A citation analysis. Journal of Documentation, 35, 179–196.
Hunt, R.K. (1990). Journal deselection in a biomedical research library: A mediated mathematical approach. Bulletin of Medical Library Association 78(1), 45–48.
Jiménez-Contreras, E., Moneda, M., De La, Olvera, D., & Ruiz De Osma Delatas, E. (1994). Determinación de las colecciones básicas de publicaciones periódicas en hemerotecas científicas [Identification of core collections of journals at academic libraries]. In IV Jornadas Españolas de Documentación Automatizada (pp. 215–221). Oviedo: Universidad.
Katz, B., & Katz, L.S. (1997). Magazines for libraries: For the general reader and school, junior college, college, university and public libraries (9th ed.). London/New York: Bowker-Saur.
Lancaster, F.W. (1993). If you want to evaluate your library (2nd ed.). London: Library Association.
Rasch, G.E. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: The Chicago University Press.
Ruiz-Baños, R. (1994). Un modelo de expurgo para las revistas científicas en bibliotecas uníversitarias [A model for weeding scientific periodicals in university libraries]. In VIII Jornadas Bibliotecarias de Andalucía (pp. 278– 288). Huelva, Spain: Diputación Provincial.
Ruiz-Baños, R., & Jiménez-Contreras, E. (1996). Envejecimiento de la literatura científica en documentación. Influencia del origen nacional de las revistas. Estudio de una muestra [Obsolescence of scientific literature in information science. Influence of the national origin of the journals. A sample study]. Revista Española de Documentación Científica 19, 39–49.
Segal, J.A. (1986). Journal deselection: A literature review and an application. Science and Techology Libraries, 6(3), 25–42.
Slote, S.J. (1997). Weeding library collections: Library weeding method (4th ed.). Englewood: Libraries Unlimited.
Terrada, M.L., Cueva, A., & Añon, R. (1979). La obsolescencia de la información científica en las publicaciones médicas españolas [Obsolescence in medical publications in Spain]. Revista Española de Documentación Científica, 2, 9–32.
Zhao, H., & Jiang, G. (1985). Shifting of world's center and scientists' social ages. Scientometrics, 8(1), 59–80.

DOI: 10.1002/asi